Spatial Statistics

Parametric spatial covariance models in the ensemble Kalman filter

Jacob Skauvold , Jo Eidsvik

Department of Mathematical Sciences, NTNU, NO-7491 Trondheim, Norway

A B S T R A C T

Several applications rely on data assimilation methods for complex spatio-temporal problems. The focus of this paper is on ensemble-based methods, where some approaches require estimation of covariances between state variables and observations in the as-similation step. Spurious correlations present a challenge in such cases as they can degrade the quality of the ensemble represen-tation of probability distributions. In particular, prediction vari-ability is often underestimated. We propose to replace the sample covariance estimate by a parametric approach using maximum likelihood estimation for a small number of parameters in a spatial covariance model. Parametric covariance and precision estimation are employed in the context of the ensemble Kalman filter, and applied to a Gauss-linear autoregressive model and a geological process model. We learn that parametric approaches reduce the underestimation in prediction variability. Furthermore rich, non-stationary models do not seem to add much over simpler models with fewer parameters.

1. Introduction

The ensemble Kalman filter (EnKF) is a popular Monte Carlo method for sequential data assimilation in complex systems (Evensen, 2009a). At each step of this approach, Monte Carlo samples, also called ensemble members, are first forecasted using the forward model and then updated with respect to data. The update step of the EnKF is based on covariances between forecast variables and data, the updated ensemble members being linear combinations of the forecast ensemble members

with weights determined by estimated covariances. Empirical covariance matrices are typically used to specify the Kalman gain, i.e. the matrix of update weights. Although this empirical approach gives unbiased estimates of covariances, the formulation tends to produce inaccurate state estimates, especially when the number of state variables is much larger than the Monte Carlo sample size (Furrer and Bengtsson, 2007; Evensen, 2009b). The effect is undesired overfitting, and ensemble representations produced by the standard EnKF typically underestimate variability (Sætrom and Omre, 2013).

Localization and inflation of the covariance are common remedies for reducing the underestimation of variance in the EnKF (Furrer and Bengtsson, 2007; Asch et al., 2016). Hierarchical Bayes formulations have also been considered as a means of stabilizing the EnKF matrix expressions (Myrseth and Omre, 2010; Ueno and Nakamura, 2016; Tsyrulnikov and Rakitko, 2017; Stroud et al., 2018). In a similar vein, penalization of the inverse covariance matrix has been used in various ways, for instance by imposing a sparse neighborhood structure (Ueno and Tsuchiya, 2009) or by an ℓ_1 norm penalty to get a sparse graph structure (Hou et al., 2016).

Albeit promising in many applications of the EnKF, none of the mentioned approaches make explicit use of the spatial elements seen in many application domains (Cressie and Wikle, 2011; Katzfuss et al., 2016). In this paper we advocate stronger links between spatial statistics and EnKF approaches to improve the properties of the analysis ensemble. Our focus is to use Gaussian random field models and spatial covariance functions in the specification of covariances entering in the Kalman gain. Within this framework we apply maximum likelihood estimation to specify covariance parameters. This geostatistical approach means that only a small number of covariance parameters must be estimated on the basis of the ensemble, reducing the risk of overfitting and giving less underestimation of prediction variability. Ueno et al. (2010) used likelihood analysis within the EnKF for estimation of parameters in the measurement model. Similarly, Ueno and Nakamura (2016) and Stroud et al. (2018) used a Bayesian formulation for parameter estimation. Our approach is different in that we embed the forecast ensemble in a Gaussian process framework, and estimate the parameters of that approximation to the forecast distribution.

Many applications for which the EnKF has turned out to be useful are characterized by complex dynamical behavior giving rise to non-stationarity. Irregular data sampling design can also lead to non-stationarity because some regions are densely sampled while others are hardly informed by data at all. A parametric approach must accommodate these aspects in a realistic manner, and we explore how a trade off between model flexibility and complexity is sought.

In Section 2 we describe the ideas underlying linear updating of an ensemble in a static situation. In Section 3 we extend this to a dynamic state-space model, using parametric covariance or precision matrices in the EnKF update. In Sections 4 and 5 we study the performance of the suggested approaches on a linear model and on an example from geology.

2. Approximate linear posterior sampling

Here we describe the underlying idea of posterior sampling with linear conditioning to data, considering a static situation. The time-dependent case is studied in Section 3.

2.1. Notation and assumptions

Let $\boldsymbol{x} = (x_1, \ldots, x_n)'$ denote the uncertain variables of interest, and $p(\boldsymbol{x})$ the prior probability density function of $\boldsymbol{x}$. The size n of the target vector is in our case equal to the number of grid cells in a discretized spatial domain, typically in the order of 10^5 or higher. We assume, as is often the case for numerical simulations of physical systems, that it is comparatively easy to generate samples from $p(\boldsymbol{x})$, but that density evaluation is difficult or infeasible. The prior ensemble consists of B independent, equally likely realizations

$$\{\boldsymbol{x}^{1,f}, \ldots, \boldsymbol{x}^{B,f}\}. \tag{1}$$

The superscript f denotes *forecast* in this context. In applications involving computer-intensive numerical simulations, the ensemble size B is usually on the order of 10 to 100 because of limitations in computing resources (processing and memory).

The data are denoted by $\boldsymbol{y} = (y_1, \ldots, y_m)'$. In our context we use the common linear and Gaussian likelihood model; $p(\boldsymbol{y}|\boldsymbol{x}) = \text{Normal}(\boldsymbol{G}\boldsymbol{x}, \boldsymbol{T})$. The posterior distribution of the target vector $\boldsymbol{x}$, given the data $\boldsymbol{y}$, is defined by $p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})$. We consider approaches that construct a posterior ensemble

$$\{\boldsymbol{x}^{1,a}, \ldots, \boldsymbol{x}^{B,a}\}, \tag{2}$$

of equally weighted realizations, approximately representative of the posterior $p(\boldsymbol{x}|\boldsymbol{y})$. The superscript a denotes *analysis* or *assimilated*.

2.2. Simulation and linear updating

When the prior is represented by a forecast ensemble (1) it is possible, in principle, to use the likelihood $p(\boldsymbol{y}|\boldsymbol{x}^{b,f})$ to re-weight the prior samples and get a posterior representation. However, methods going in this direction, such as the particle filter (Doucet et al., 2000), tend to place all weight on one ensemble member in high-dimensional settings (Snyder et al., 2008). Hence, practical use of these approaches is limited. One can try to reduce the data dimension in various ways, for instance by conditioning only on some summary of the data as in approximate Bayesian computation (Beaumont, 2010), but methods of this type typically require that a large number of proposed realizations be generated, to the point of having a prohibitive computational cost in the kind of setting we are envisioning. Nor is it clear how to construct viable summary statistics or acceptance criteria for large spatial models.

We focus on approaches that use linear updating to construct the analysis ensemble (2). This class of updating schemes correctly approximates the posterior distribution when the prior distribution and likelihood are both Gaussian, and when the ensemble size tends to infinity. While no performance guarantees can be given in the general case, for instance when assumptions of Gaussianity cannot be justified, this approach has shown itself to be very useful in several applications (Asch et al., 2016).

The linear update means that approximate posterior samples are generated by

$$\boldsymbol{x}^{b,a} = \boldsymbol{x}^{b,f} + \hat{\boldsymbol{K}}(\boldsymbol{y} - \boldsymbol{y}^b), \tag{3}$$

where $\hat{\boldsymbol{K}}$ is a weight matrix or gain that must be specified, and $\boldsymbol{y}^b$ is a synthetic observation or perturbed model equivalent given by

$$\boldsymbol{y}^b = \boldsymbol{G}\boldsymbol{x}^{b,f} + \boldsymbol{\epsilon}^b, \tag{4}$$

and the observation error realization $\boldsymbol{\epsilon}^b$ is drawn from a zero-mean multivariate normal with covariance matrix $\boldsymbol{T}$. Underlying the update in (3) is the joint covariance

$$\text{Cov}\left(\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix}\right) = \begin{bmatrix}\boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{x,y}\\ \boldsymbol{\Sigma}_{y,x} & \boldsymbol{\Sigma}_y\end{bmatrix}, \tag{5}$$

from which the gain matrix $\boldsymbol{K}$ is defined as

$$\boldsymbol{K} = \boldsymbol{\Sigma}_{x,y}\boldsymbol{\Sigma}_y^{-1}. \tag{6}$$

When the model is correctly specified, the gain matrix in (6) is the optimal linear regression weight for regressing the forecast state ensemble in (3) on the ensemble of synthetic observations in (4). In practice, the optimal gain is unknown, and an estimated gain matrix $\hat{\boldsymbol{K}}$ is obtained from samples $\boldsymbol{x}^{b,f}$ and $\boldsymbol{y}^b$, $b = 1, \ldots, B$. Since we assume that the likelihood model, including $\boldsymbol{G}$ and $\boldsymbol{T}$, is known, the ensemble is only used to estimate the prior covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_x$. The estimated gain then becomes

$$\hat{\boldsymbol{K}} = \hat{\boldsymbol{\Sigma}}_{x,y}\hat{\boldsymbol{\Sigma}}_y^{-1} = \hat{\boldsymbol{\Sigma}}\boldsymbol{G}'(\boldsymbol{G}\hat{\boldsymbol{\Sigma}}\boldsymbol{G}' + \boldsymbol{T})^{-1}. \tag{7}$$

We will also use of a formulation with the precision matrix $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$, which sometimes has a sparse (Markovian) structure. Moreover, for some models one can incorporate non-stationarity

directly through the precision structure (Section 3.3). An algebraically equivalent formulation of the assimilation step specified by (3), (4) and (6) is then

$$\left[\hat{\boldsymbol{Q}} + \boldsymbol{G}'\boldsymbol{T}^{-1}\boldsymbol{G}\right](\boldsymbol{x}^{b,a} - \boldsymbol{x}^{b,f}) = \boldsymbol{G}'\boldsymbol{T}^{-1}(\boldsymbol{y} - \boldsymbol{y}^{b}). \tag{8}$$

Computing the updated ensemble using this expression requires the solution of a system of linear equations with coefficient matrices that are sparse in most cases.

2.3. Empirical and parametric covariance specification

An empirical or non-parametric estimate of the covariance based on the prior ensemble is

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{B}\sum_{b=1}^{B}(\boldsymbol{x}^{b,f} - \bar{\boldsymbol{x}}^{f})(\boldsymbol{x}^{b,f} - \bar{\boldsymbol{x}}^{f})', \quad \bar{\boldsymbol{x}}^{f} = \frac{1}{B}\sum_{b=1}^{B}\boldsymbol{x}^{b,f}. \tag{9}$$

When the dimension of $\boldsymbol{x}$ is large, compared with the sample size B, the empirical estimates of the sample covariances are prone to large Monte Carlo errors (Furrer and Bengtsson, 2007; Sætrom and Omre, 2011).

In a parametric approach, the covariance is defined by a few model parameters $\boldsymbol{\theta}$, and we use $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ to denote the covariance matrix controlled by this parameter vector. The parameters must be chosen so that the resulting covariance matrix describes the simulation results well. Using a likelihood function

$$l(\boldsymbol{\theta}) = l(\boldsymbol{\theta}; \boldsymbol{x}^{1,f}, \ldots, \boldsymbol{x}^{B,f}) \tag{10}$$

for this purpose, the parameter estimate is

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} l(\boldsymbol{\theta}). \tag{11}$$

We assume that the likelihood is representative of a Gaussian process, where the mean is computed directly from the ensemble. Moreover, the B ensemble members are assumed to be independent and identically distributed, so that the likelihood is given by

$$l(\boldsymbol{\theta}) = -\frac{B}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}| - \frac{1}{2}\sum_{b=1}^{B}(\boldsymbol{x}^{b,f} - \bar{\boldsymbol{x}}^{f})'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x}^{b,f} - \bar{\boldsymbol{x}}^{f}). \tag{12}$$

The parametric estimate of the covariance matrix $\boldsymbol{\Sigma}$ is then $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$.

For common parametrizations of spatial dependence in $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, there are closed form expressions for the derivatives of the likelihood (12), see e.g. Petersen et al. (2008). These are calculated at every iteration of the optimization procedure. Parameter estimates typically converge after no more than 5 to 10 Fisher-scoring iterations,

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} + \left[E\left(\frac{d^2 l(\hat{\boldsymbol{\theta}})}{d\boldsymbol{\theta}^2}\right)\right]^{-1}\frac{dl(\hat{\boldsymbol{\theta}})}{d\boldsymbol{\theta}}. \tag{13}$$

If derivatives are not available, other optimization schemes must be used, such as Nelder–Mead search (Lagarias et al., 1998).

2.4. Illustrative spatial example

We compare the results of empirical and parametric covariance estimates for different sample sizes ($B = 100$ and $B = 1000$). The spatial variable $\boldsymbol{x}$ is here represented on a regular 25×25 grid, and entry x_i represents the variable in grid cell $i \in \{1, \ldots, 625\}$. The prior mean values are 0, and the covariance model is stationary with variance $\sigma^2 = 1$ in all grid cells and an exponential correlation function. Defining $\boldsymbol{D}$ to be the 625×625 matrix of distances between all grid cells, the true covariance matrix is $\boldsymbol{\Sigma} = \sigma^2 \exp(-3\boldsymbol{D}/\eta)$, where $\eta = 10$ indicates an effective correlation range of 10 grid cells.

Table 1
Performance of different covariance specifications. Matrix norms are means of 100 replicates, and parentheses represent the standard deviation of these.

$B = 100$	Kullback–Leibler	Bhattacharyya	Frobenius
emp	96 000 (741)	558 (0.212)	63.3 (2.6)
emp, loc	113 (1.6)	26.2 (0.3)	88.1 (0.2)
par	0.01 (0.01)	0.002 (0.002)	4.54 (3.2)
semi-par	11.5 (0.9)	2.78 (0.2)	31.1 (1.4)
$B = 1000$	Kullback–Leibler	Bhattacharyya	Frobenius
emp	370 (2.7)	40.3 (0.2)	20.2 (0.8)
emp, loc	52.1 (0.07)	16 (0.07)	87 (0.02)
par	0.001 (0.001)	0.0004 (0.0003)	1.8 (1.3)
semi-par	5.6 (0.5)	1.36 (0.1)	13.6 (1.6)

We study covariance estimation performance using three criteria: Kullback–Leibler divergence, Bhattacharyya distance and the Frobenius norm. For all these measures we compare the specified covariance $\hat{\boldsymbol{\Sigma}}$ with the true covariance matrix $\boldsymbol{\Sigma}$. For zero-mean multivariate Gaussian vectors, the Kullback–Leibler divergence D_{KL}, Bhattacharyya distance D_{B} and Frobenius norm distance D_{F} between $\hat{\boldsymbol{\Sigma}}$ and $\boldsymbol{\Sigma}$ are

$$D_{\mathrm{KL}}(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \frac{1}{2}[\mathrm{trace}(\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}) - n + \log|\hat{\boldsymbol{\Sigma}}| - \log|\boldsymbol{\Sigma}|], \tag{14}$$

$$D_{\mathrm{B}}(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \frac{1}{2}\log|\tilde{\boldsymbol{\Sigma}}| - \frac{1}{4}|\log|\hat{\boldsymbol{\Sigma}}| - \frac{1}{4}\log|\boldsymbol{\Sigma}|, \quad \tilde{\boldsymbol{\Sigma}} = [\boldsymbol{\Sigma} + \hat{\boldsymbol{\Sigma}}]/2, \tag{15}$$

$$D_{\mathrm{F}}(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \sqrt{\mathrm{trace}[(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})']}. \tag{16}$$

Results of the covariance estimation are presented in Table 1. The empirical estimate (emp) is poor for all measures, even though the norms decrease when the sample size B is increased. When using a localized version of the empirical approach (emp, loc), the performance is clearly improved from the straightforward empirical covariance specification method, except for the Frobenius norm which might carry some localization artifacts. Localization is here done by elementwise multiplication of the estimated covariance matrix with a tapering matrix setting covariance entries beyond a 10-cell range to 0. In Table 1 we further see that the parametric approach (par), which has the same form as the generating mechanism in this case, is clearly the best for all norms. For a semi-parametric approach (semi-par), the norms are smaller than for the localized empirical approach. In the semi-parametric approach we set the diagonal entries of the covariance matrix from sample variances, while a single exponential correlation decay parameter is estimated by maximizing the likelihood, given the assigned variances. (See Section 3.) In summary, the results indicate that the empirical approaches have very large Monte Carlo errors. They do not estimate $\boldsymbol{\Sigma}$ very well.

We next simulate data to study properties of the different covariance specification approaches under linear data updating. Data are collected at all locations in the 25×25 grid, according to $y(\boldsymbol{s}_i) = x(\boldsymbol{s}_i) + N(0, \tau^2)$, $i = 1, \ldots, m = n = 625$, with $\tau = 0.5$. As every grid cell is observed, the conditioning will not introduce significant non-stationarity, and apart from edge effects the posterior covariance will be stationary. We study the performance in terms of posterior mean square prediction error (MSPE), ensemble coverage probabilities (CovPr) at the 80% nominal level and continuously ranked probability score (CRPS), see e.g. Gneiting et al. (2007).

Table 2 summarizes the results. For all prediction measures, the fully parametric and semi-parametric approaches give the best results. The localized empirical method is a little worse, but much better than the straightforward empirical estimate. When the sample size increases, the latter improves markedly, but for $B = 1000$ it is still not at the performance levels of the other approaches. The poor performance of the empirical approach is largely due to sampling variability causing erroneous covariances which are again influencing the linear updating. In particular, the coverage probabilities of this empirical approach are very small at the 80% nominal level.

Table 2
Performance of linear updating using different covariance specifications. The results are means of 500 replicates, and the parentheses represent the standard deviation of these.

$B = 100$	MSPE	CovPr(80)	CRPS
emp	0.372 (0.035)	32.3 (2.1)	0.241 (0.007)
emp, loc	0.263 (0.009)	76.7 (1.8)	0.209 (0.007)
par	0.245 (0.008)	79.2 (1.9)	0.197 (0.006)
semi-par	0.246 (0.008)	79.1 (1.9)	0.198 (0.006)
$B = 1000$	MSPE	CovPr(80)	CRPS
emp	0.252 (0.011)	72.7 (1.9)	0.216 (0.007)
emp, loc	0.2622 (0.008)	79.7 (1.7)	0.205 (0.006)
par	0.2470 (0.008)	80.0 (1.6)	0.198 (0.006)
semi-par	0.2471 (0.007)	80.0 (1.6)	0.198 (0.006)

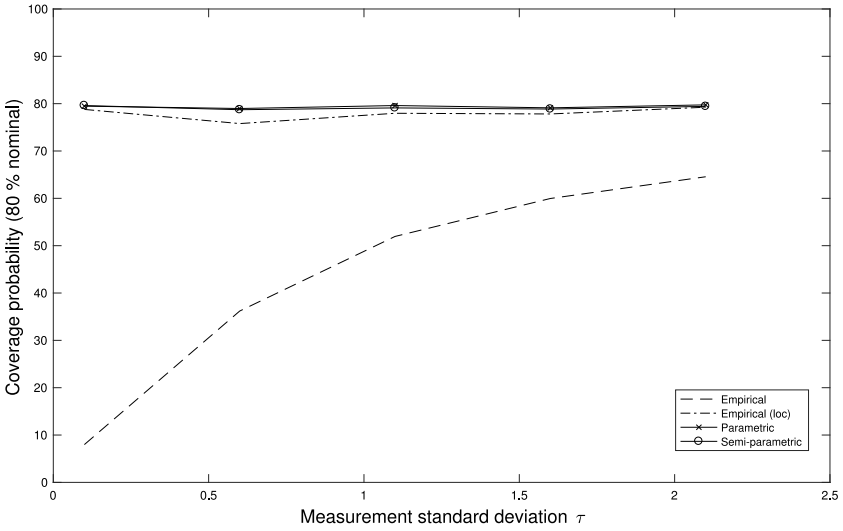

Fig. 1. Coverage probabilities for different measurement noise levels in the data.

Fig. 1 shows the relationship between the coverage probabilities and the measurement noise standard deviation τ.

The two parametric approaches are close to the nominal level of 80% for all noise levels, for $B = 100$. The localized empirical specification also performs well, while the straightforward empirical approach is very poor for small noise levels and only gradually goes towards the nominal level for larger noise levels. It is not surprising that the deviation from the nominal coverage level gets smaller for large measurement errors since the data has little influence on the update in that case. The very low coverage (10%) for noise standard deviation 0.1 is more surprising. In fact, one might expect the gain $\boldsymbol{\Sigma}(\boldsymbol{\Sigma} + \tau^2\boldsymbol{I})^{-1}$ to be close to the identity matrix, because the addition of τ^2 is negligible. In this case, however, the Monte Carlo errors in the sample covariance matrix are too large relative to τ^2.

Since the computing time is larger for the parametric approaches one could argue that for a fair comparison a larger sample size should be used for the empirical approaches. Then again, typical applications of linear updating have long evaluation times for the mechanism providing $\boldsymbol{x}^{b,f}$, $b = 1, \ldots, B$, so the additional time spent on covariance estimation is negligible in comparison.

3. Parametric covariance estimation in the EnKF

Here we extend the parametric covariance estimation approach to dynamical systems. We begin by presenting some modeling and methodological assumptions. Then we describe the updating scheme incorporating parametric covariance estimation. Finally we show how this scheme can be applied to non-stationary models.

3.1. Assumptions

The state vector is denoted $\boldsymbol{x}_t = (x_{t,1}, \ldots, x_{t,n})'$, for time $t = 0, 1, \ldots, N$. Assuming a prior density $p(\boldsymbol{x}_0)$ at the initial step, the state evolves according to a dynamic model

$$\boldsymbol{x}_t = \boldsymbol{f}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\delta}_t), \quad t = 1, \ldots, N, \tag{17}$$

where $\boldsymbol{\delta}_t$ is a noise term and the functional relationship defined by $\boldsymbol{f}_t$ is known. In realistic situations, this relationship is often obtained by forward integration of a system of differential equations. Moreover, data at time $t = 1, \ldots, N$ is denoted by $\boldsymbol{y}_t = (y_{t,1}, \ldots, y_{t,m})'$ and the likelihood model is defined by

$$\boldsymbol{y}_t = \boldsymbol{G}_t\boldsymbol{x}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \text{Normal}(\boldsymbol{0}, \boldsymbol{T}_t), \quad t = 1, \ldots, N, \tag{18}$$

where the design matrix $\boldsymbol{G}_t$ and covariance matrix $\boldsymbol{T}_t$ are known.

The goal of filtering is assessing the conditional density $p(\boldsymbol{x}_t|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t)$, for times $t = 1, \ldots, N$. Because of the non-linear relationship in (17), there is no closed form expression for the filtering density. The EnKF sequentially computes and maintains an ensemble representation of the filtering distribution at all times. Assimilation is accomplished by linear updates of ensemble members with respect to observations. Starting from an analysis ensemble representation $\boldsymbol{x}_{t-1}^{b,a}$, $b = 1, \ldots, B$, at the previous time step, the EnKF iteration proceeds in two steps:

(i) **Forecasting** by advancing each ensemble member through time by forward integration of the dynamical model,

$$\boldsymbol{x}_t^{b,f} = \boldsymbol{f}(\boldsymbol{x}_{t-1}^{b,a}, \boldsymbol{\delta}_t^b), \quad b = 1, \ldots, B. \tag{19}$$

(ii) **Updating** the ensemble members with respect to data, based on a linear relationship between the two,

$$\boldsymbol{x}_t^{b,a} = \boldsymbol{x}_t^{b,f} + \hat{\boldsymbol{K}}_t^{-1}(\boldsymbol{y}_t - \boldsymbol{y}_t^{b,f}) \tag{20}$$

$$\hat{\boldsymbol{K}}_t = \hat{\boldsymbol{\Sigma}}_t\boldsymbol{G}_t'(\boldsymbol{G}_t\hat{\boldsymbol{\Sigma}}_t\boldsymbol{G}_t' + \boldsymbol{T}_t)^{-1}. \tag{21}$$

As in the static case described in Section 2, the Kalman gain relies on an estimate of the forecast covariance matrix $\boldsymbol{\Sigma}_t = \text{Cov}(\boldsymbol{x}_t|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1})$. The standard formulation of the EnKF uses the empirical or non-parametric covariance matrix of the forecast ensemble for this purpose,

$$\hat{\boldsymbol{\Sigma}}_t = \frac{1}{B}\sum_{b=1}^{B}(\boldsymbol{x}_t^{b,f} - \bar{\boldsymbol{x}}_t^f)(\boldsymbol{x}_t^{b,f} - \bar{\boldsymbol{x}}_t^f)', \qquad \bar{\boldsymbol{x}}_t^f = \frac{1}{B}\sum_{b=1}^{B}\boldsymbol{x}_t^{b,f}. \tag{22}$$

However, as was discussed in the previous section, this direct empirical estimate is prone to large Monte Carlo errors. We proceed instead by describing parametric estimates of this covariance matrix, or the associated precision matrix.

3.2. Parametric EnKF update

Denote a parametric specification of the forecast covariance by $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t,\boldsymbol{\theta}_t}$. As suggested in (11) and (12), a parametric specification of the covariance is defined by $\boldsymbol{\Sigma}_{t,\hat{\boldsymbol{\theta}}_t}$ where

$$\hat{\boldsymbol{\theta}}_t = \underset{\boldsymbol{\theta}_t}{\text{argmax}}\, l(\boldsymbol{\theta}_t; \boldsymbol{x}_t^{1,f}, \ldots, \boldsymbol{x}_t^{B,f}). \tag{23}$$

Again we assume this likelihood is that of a Gaussian process and, although there will be coupling over time because the Kalman gain is formed from ensemble-based covariance estimates, we proceed as if the B ensemble members are independent and identically distributed. This means that the likelihood is

$$l(\boldsymbol{\theta}_t) = -\frac{B}{2}\log|\boldsymbol{\Sigma}_{t,\boldsymbol{\theta}_t}| - \frac{1}{2}\sum_{b=1}^{B}(\boldsymbol{x}_t^{b,f} - \bar{\boldsymbol{x}}_t^f)'\boldsymbol{\Sigma}_{t,\boldsymbol{\theta}_t}^{-1}(\boldsymbol{x}_t^{b,f} - \bar{\boldsymbol{x}}_t^f). \tag{24}$$

In our formulation the parameters will vary over time. However, the change from one time point to the next tends to be small, so we start the optimization of the likelihood using the estimate from the previous time step. The actual maximization will depend on the functional form of the parametric covariance model, and whether derivatives are available (see (13)).

Again, it can sometimes be useful to fit parameters of the precision matrix $\boldsymbol{Q}_{t,\boldsymbol{\theta}_t}$, rather than working with the covariance matrix (see Section 3.3).

3.3. Choice of parametric models

Common spatial covariance functions include the spherical and Matern-type with the exponential and the Gaussian as extreme cases (Cressie and Wikle, 2011). The exponential covariance function was used in Section 2.4.

In the simplest, stationary model, the forecast variances $\text{Diag}(\boldsymbol{\Sigma}_{t,\boldsymbol{\theta}_t})$, are the same at all spatial locations, and pairwise correlation depends only on distance and not on specific locations. This is attractive from a computational point of view because there are only a few model parameters to estimate. For instance, the exponential covariance function has one variance parameter and one correlation decay parameter. Assuming that the target random field is stationary might be unrealistic in situations where the dynamical model affects various parts of the domain differently. Also, sparse data would lead to much smaller variance near data locations than far away, and possibly to a non-stationary correlation decay.

Non-stationary models are more flexible and, in the context of data assimilation for non-linear dynamical systems, arguably better suited at capturing relevant features of the spatio-temporal process. The main challenge of building a non-stationary model is that there are several kinds of non-stationarity. Which kind is useful for our situation? We discuss some approaches, and then pursue a couple of them in more detail.

There are several popular non-stationary covariance models, e.g. Paciorek and Schervish (2006) and Jun and Stein (2008). Various attempts have been made to impose structure in the spatial domain or via spatially varying covariates (Neto et al., 2014; Parker et al., 2016). However, non-stationary spatial models have been found to be relatively difficult to parameterize, mainly due to the requirement that the fitted model must give positive definite covariance matrices for any configuration of spatial sites. Another approach involves non-stationary modeling of the precision matrix or inverse covariance matrix: Fuglstad et al. (2015a) used spatially dependent basis functions to represent the precision structure. However, it has been difficult to estimate model parameters in such rich model formulations, particularly when many basis functions are involved, and sometimes much simpler parsimonious models perform equally well in practice (Fuglstad et al., 2015b).

The first non-stationary model we consider here is a semi-parametric approach where marginal variances can differ between spatial locations, while the correlation structure is the same everywhere (semi-par in Section 2.4), see also Asfaw and Omre (2016). This model entails that the diagonal entries $\hat{\sigma}_{t,1}^2, \ldots, \hat{\sigma}_{t,n}^2$ of the forecast covariance matrix $\hat{\boldsymbol{\Sigma}}_t$ are specified empirically from the data,

$$\hat{\sigma}_{t,i}^2 = \frac{1}{B}\sum_{b=1}^{B}(x_{t,i}^{f,b} - \bar{x}_{t,i}^f)^2. \tag{25}$$

Assuming a parametric spatial correlation function, the likelihood is maximized using fixed variances as calculated in (25), meaning the likelihood is

$$l(\boldsymbol{\theta}_t) = -\frac{B}{2}\log|\boldsymbol{\Sigma}_{t,\boldsymbol{\theta}_t}| - \frac{1}{2}\sum_{b=1}^{B}(\boldsymbol{x}_t^{b,f} - \bar{\boldsymbol{x}}_t^f)'\boldsymbol{\Sigma}_{t,\boldsymbol{\theta}_t}^{-1}(\boldsymbol{x}_t^{b,f} - \bar{\boldsymbol{x}}_t^f),$$

$$\boldsymbol{\Sigma}_{t,\boldsymbol{\theta}_t} = \text{diag}(\hat{\sigma}_{t,1}, \ldots, \hat{\sigma}_{t,n})\boldsymbol{R}_{t,\boldsymbol{\theta}_t}\text{diag}(\hat{\sigma}_{t,1}, \ldots, \hat{\sigma}_{t,n}), \tag{26}$$

where $\boldsymbol{R}_{t,\boldsymbol{\theta}_t}$ is the correlation matrix with unknown parameters, and diag$(\cdot)$ forms a diagonal matrix of the vector input.

Another parametric model we consider here is based on a stochastic partial differential equation (SPDE) formulation (Lindgren et al., 2011). Let Δ be the Laplacian operator, κ a model parameter and $\mathcal{W}(\mathbf{s})$ a white noise spatial process, and define the process $x(\boldsymbol{s})$ by

$$(\kappa^2 - \Delta)^{\alpha/2}x(\mathbf{s}) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^2. \tag{27}$$

Lindgren et al. (2011) showed how spatial discretization naturally connects the SPDE in (27) to the precision matrix coefficients of the Gaussian Markov random field representation. The model parameters κ and α are not directly interpretable like the marginal variance or correlation decay, but Lindgren et al. (2011) showed closed-form relations between these parameters and variance, correlation range and smoothness parameters for the Matern covariance function. Fuglstad et al. (2015a) extended the SPDE in (27) to a non-stationary formulation,

$$(\kappa^2(\boldsymbol{s}) - \nabla \cdot \boldsymbol{H}(\boldsymbol{s})\nabla)^{\alpha/2}x(\mathbf{s}) = \mathcal{W}(\mathbf{s}). \tag{28}$$

Here, ∇ is the differential operator and the 2×2 matrix $\boldsymbol{H}(\boldsymbol{s})$ contains basis functions with location-dependent covariates, giving non-stationarity. We present a particular parametrization for the geological example in Section 5, where the shape of a basis function is set on the basis of information from the forecast ensemble.

With precision matrix $\boldsymbol{Q}_{t,\boldsymbol{\theta}_t}$, the likelihood equals

$$\ell(\boldsymbol{\theta}_t) = \frac{B}{2}\log|\boldsymbol{Q}_{t,\boldsymbol{\theta}_t}| - \frac{1}{2}\sum_{b=1}^{B}(\boldsymbol{x}_t^{b,f} - \bar{\boldsymbol{x}}_t^f)^T\boldsymbol{Q}_{t,\boldsymbol{\theta}_t}(\boldsymbol{x}_t^{b,f} - \bar{\boldsymbol{x}}_t^f). \tag{29}$$

The maximum likelihood estimate of parameter $\boldsymbol{\theta}_t$ is computed by optimizing (29). As stated in Fuglstad et al. (2015a), the parametrization should not be too rich, as the optimization procedure can be hampered by a difficult likelihood surface. Analytical expressions for log-likelihood derivatives are available in some cases. However, for stability reasons, the numerical experiments of this paper use the simplex method for derivative-free optimization.

4. Simulation study for linear dynamic model

In this part we extend the 25×25 grid example to an autoregressive case in the spatio-temporal domain (Cressie and Wikle, 2011). As in Section 2.4, the Gaussian initial distribution has mean 0, a covariance matrix $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}$ specifying a variance of 1 and an exponential covariance function with effective correlation range $\eta = 10$ cells. The dynamic model is

$$x_t(\mathbf{s}) = \phi x_{t-1}(\mathbf{s}) + \delta_t(\mathbf{s}), \quad \boldsymbol{\delta}_t \sim N(\mathbf{0}, (1-\phi^2)\boldsymbol{\Sigma}), \quad t = 1, \ldots, 10,$$

for all grid cells $\boldsymbol{s}$. With this covariance structure for the additive noise terms, $\boldsymbol{x}_t$ is a stationary spatial process over time. In the experiments we set the autoregressive coefficient $\phi = 0.9$.

The data gathering scheme is defined by sampling at $m = 15$ irregular sites, which are the same at all 10 time steps. This sparse sampling scheme will induce non-stationarity in the covariance over time. The measurement noise terms are independent with variance 0.5^2.

Estimation approaches are again compared in terms of MSPE, coverage probability and CRPS. We consider two locations: Grid cell (2,13), which is far from data locations, and grid cell (18,13) which is near data locations. Results for ensemble size $B = 100$ are shown in Table 3. These are averages over 500 replicates.

We notice that the straightforward empirical or non-parametric EnKF approach (emp) underestimates the variability in the prediction, especially for the cell far from data. Localization (emp, loc) improves this coverage problem, but it seems to give larger MSPE than the other approaches for sites far from data. The performance of the localized approach might be improved by tuning the tapering matrix, but considering the results of a particular taper still gives a basis for comparison with the other

Table 3
Performance of EnKF variants using $B = 100$ ensemble members. Results at time step 10 in the autoregressive process.

	Far from data — Cell (2,13)			Near data — Cell (18,13)		
	MSPE	CovPr(80)	CRPS	MSPE	CovPr(80)	CRPS
emp	1.93	65	0.62	0.23	74	0.20
emp, loc	1.94	79	0.55	0.23	78	0.19
par	1.89	80	0.54	0.27	78	0.21
semi-par	1.89	80	0.54	0.23	78	0.19

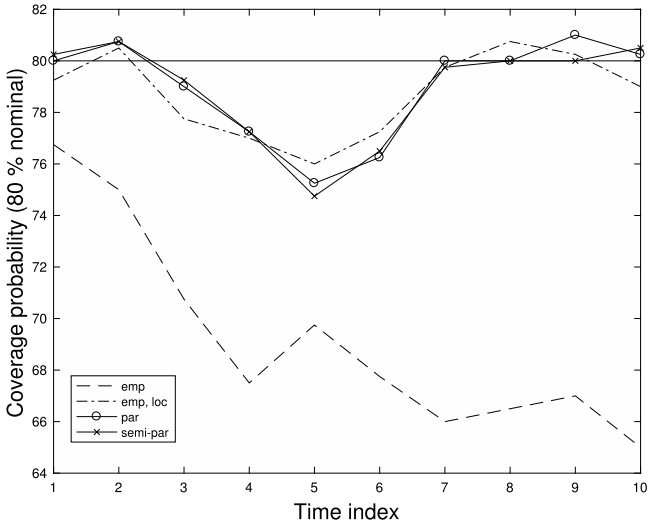

Fig. 2. Coverage probabilities for the grid cell far from data, plotted over time indices.

estimation methods. Perhaps surprisingly, the results of the simple parametric approach (par) are rather good even for this sparse design, where the true covariance is non-stationary. For the cell near data, however, its MSPE and CRPS are larger than those of the semi-parametric (semi-par) approach, which has the overall best performance.

Fig. 2 shows the coverage probabilities (at the 80% nominal level) plotted against time indices. All approaches are shown for the grid cell far from data. The probabilities are roughly constant over time, except for the empirical approach where probabilities decline, likely due to the coupling of the ensemble members in the estimation of the Kalman gain (Sætrom and Omre, 2013). This effect is much smaller with localization and with parametric specification of the covariance parameters.

For this linear and Gaussian dynamical model, the optimal solution is provided by the Kalman filter, and we next compare the filtering results at time step 10 with this solution. The Kalman filter results are shown in Fig. 3, empirical EnKF results in Fig. 4, and semi-parametric EnKF results in Fig. 5.

The prediction obtained by the semi-parametric approach shows less small-scale variability than the straightforward empirical solution, because of smaller Monte Carlo errors in the covariance estimates. This smoothness makes the semi-parametric prediction more similar to the Kalman filter result. Moreover, the empirical approach has smaller standard deviations on average, leading to the low coverage probability in Table 3.

Fig. 6 shows the exact prediction covariance (solid) of this model at time step 10. This is plotted for the grid cells near and far from data, with the horizontal axis giving distance measured towards the south from each starting point. The covariance at distance 0 is much higher for the cell far from data, but it is more difficult to detect differences in the rate of decay with distance. Along with the exact calculation, the display shows the fitted prediction covariance using semi-parametric estimation

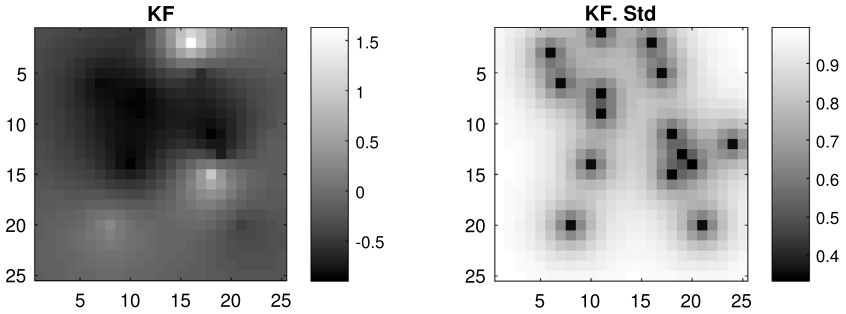

Fig. 3. Prediction (left) and prediction standard deviation (right) of the Kalman filter, at time step 10.

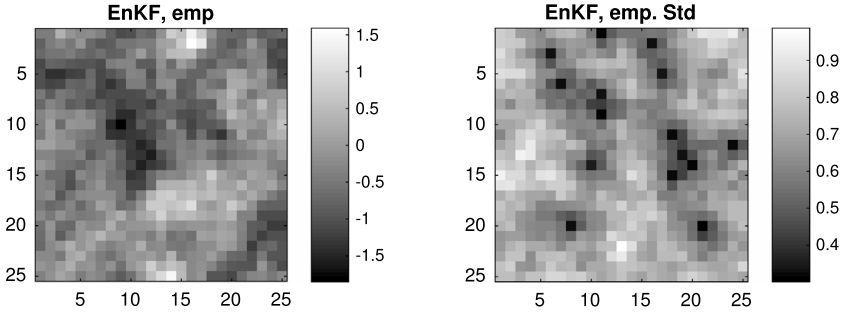

Fig. 4. Prediction (left) and prediction standard deviation (right) of the standard empirical EnKF, at time step 10.

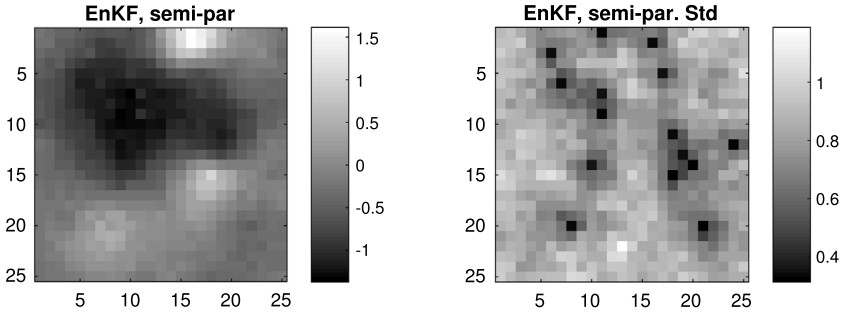

Fig. 5. Prediction (left) and prediction standard deviation (right) of the semi-parametric EnKF approach, at time step 10.

(dashed). There are random variations caused by the empirical variance estimates, but the covariance decay appears similar to the Kalman filter results, indicating that non-stationarity in correlation is moderate for this sampling design. This reasonably good fit in terms of covariance seems to account for the good prediction efficiency of the semi-parametric approach.

5. Example: Geological process model

We now apply the EnKF with parametric covariance estimation to a non-linear data assimilation problem from geology. The Geological Process Model (GPM) simulates erosion, transport and deposition of clastic sediment on length scales of tens to hundreds of kilometres over millions of years (Tetzlaff, 2005). In this case it is used to simulate sedimentation taking place in block F3 of

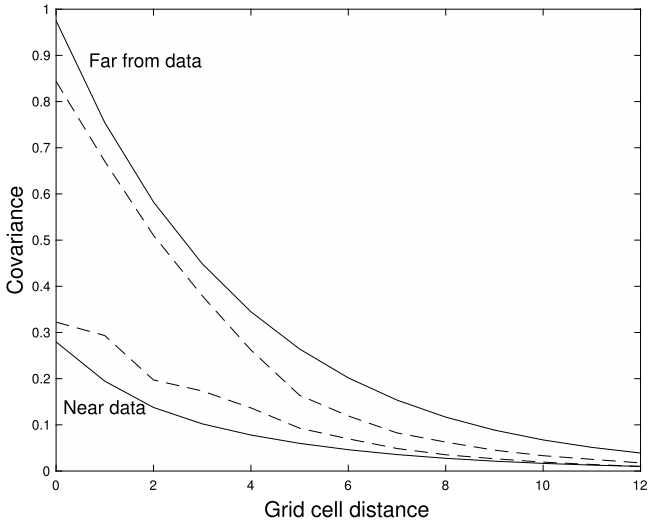

Fig. 6. Prediction covariances in the south direction from two grid cells near and far from data locations, at time step 10. Exact covariance (solid) and the fitted by the semi-parametric approach (dashed).

the Dutch sector of the North Sea between 5 and 3.5 million years ago, during the early to mid Pliocene (Ogg et al., 2016).

Statistical inversion of geological process models has been studied by e.g. Charvin et al. (2009) who used Markov chain Monte Carlo sampling and Skauvold and Eidsvik (2018) who used variants of the EnKF and the ensemble smoother. The estimation task considered here is part of a larger data assimilation problem. A more complete analysis would also track the sediment type composition or grain size distribution in the layer structure, and might additionally estimate changing environmental controls on the sedimentation process, such as sea level and sediment supply. Wingate et al. (2016) considered an event-based geological model trading off physical verisimilitude for computational speed and convenience of conditioning, and took a probabilistic programming approach coupled with variational inference to condition this model to observations.

This geological inversion task is an atypical application of the EnKF because the observations, being borehole measurements, were collected long after the sedimentation process had finished. In the region of interest, sediment age is assumed to increase monotonically with depth, i.e. younger layers are always found above older layers. Data in the form of measurements along a well trajectory can therefore be regarded as a time series, which suggests that conditioning could be carried out sequentially. Sequential conditioning procedures that halt the simulation at every time step to update the intermediate state with respect to data are attractive because they tend to be more stable and computationally efficient than approaches where the simulation is run to completion before data are taken into account. For an example of sequential conditioning of a geological model to data, see the method proposed by Parquer et al. (2017) to trace the developing morphology of meandering rivers backwards in time starting from their present-day configuration.

5.1. Problem description and setup

Fig. 7 shows the rectangular model region, which measures 66 km in the East–West direction and 37 km in the North–South direction, discretized into a regular two-dimensional grid with a 0.5 km $\times$ 0.5 km resolution, for a total of $n = 9900$ cells.

The 1.5 million year time interval covered by the simulation is discretized into 15 time steps of 100 000 years each, indexed by $t = 0, 1, \ldots, 15$. The simulator takes as input the initial elevation of the model region, i.e. the surface shown in Fig. 7. Sediment then accumulates on top of this surface

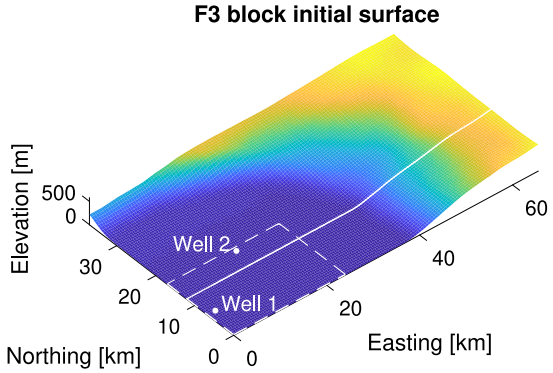

Fig. 7. Overview of the modeled region. The rectangular boundary of the F3 block is indicated by the dashed, white lines. To facilitate simulation of sediment influx the domain has been extended towards the north and east. Also included in the figure are two well locations where the thickness of the accumulating layer package is observed, and the location of the 2D section in Fig. 8, shown as a solid, white line.

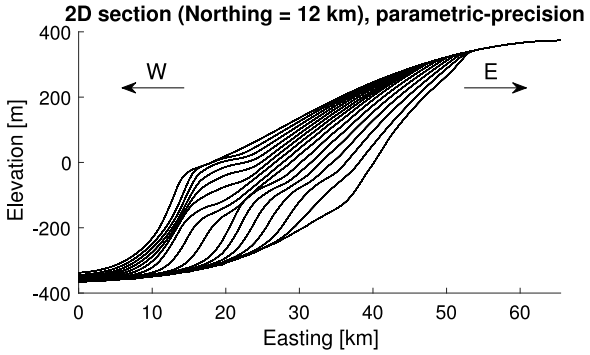

Fig. 8. Two-dimensional section through estimated layer package at $t = 15$, showing internal structure. One new layer is produced at the top of the stack every time step, leading to a strict chronological ordering with the oldest layers at the bottom. The location of the section is shown in Fig. 7 as a solid, white line.

over time, producing a layered structure whose thickness tends to increase more or less monotonically over time. Fig. 8 shows a 2D section through a simulated stack of 15 layers.

At time t the representation of the simulated layer package consists of $t+1$ layer boundary surfaces $\mathbf{z}_{kt} \in \mathbb{R}^n$, $k = 0, 1, \ldots, t$. These correspond to the black curves in Fig. 8. The cumulative thickness of all deposited sediment at time t is

$$\mathbf{x}_t = \mathbf{z}_{tt} - \mathbf{z}_{0t}, \tag{30}$$

and this will be the variable of interest here.

Noisy observations of cumulative thickness are available at every time step at two sites in the F3 block domain: Well 1 near the western boundary of the block and Well 2 near the northern boundary. These locations are shown in Fig. 7. For this example, the observations have been generated by running the simulator with known input to produce a reference realization of $\mathbf{z}_{kt}$ for $k = 1, 2, \ldots, t$ and $t = 1, 2, \ldots, N$ from which thicknesses were computed. Finally, Gaussian noise was added to the reference thickness values.

The goal of the data assimilation exercise is now to estimate the entire thickness field $\boldsymbol{x}_t$ at each time t, given the noisy measurements $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t$ at the well locations. We solve this filtering problem using three different versions of the EnKF: (a) EnKF using a semi-parametric covariance model with

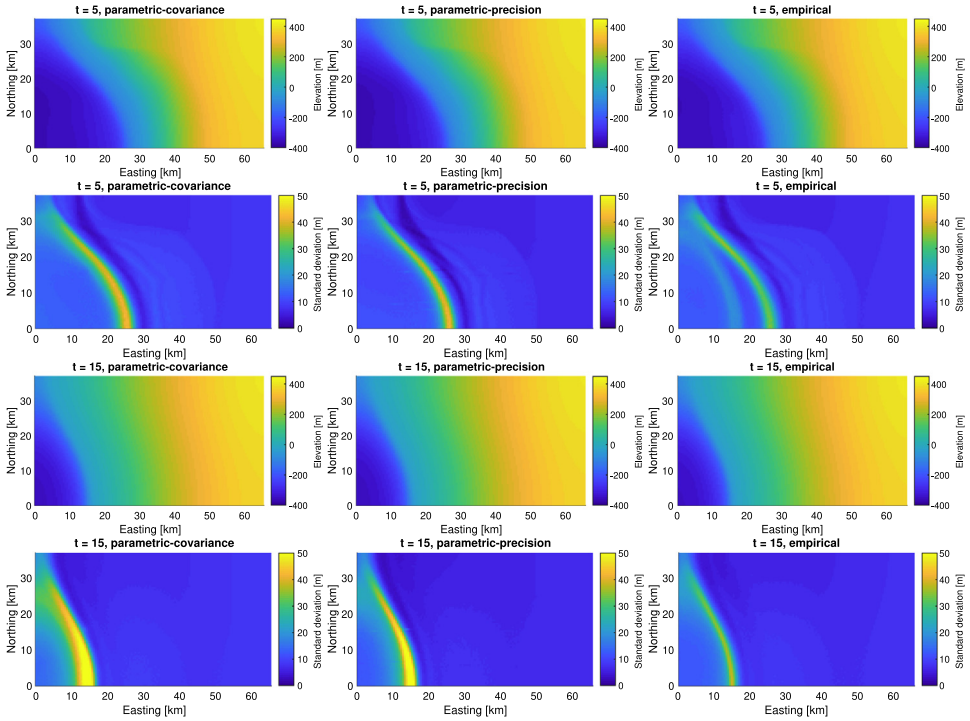

Fig. 9. Ensemble mean and standard deviation of top surface $\mathbf{z}_{tt}$ at time $t = 5$ and $t = 15$ for three EnKF variants.

empirical variances and likelihood estimation of a single correlation decay parameter. (b) EnKF with a parametric representation of the precision matrix in the SPDE representation described in Section 3.3. (c) Empirical or nonparametric standard EnKF approach. Each EnKF variant is run once with $B = 50$ ensemble members.

In the precision parametrization, a three-element parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T$ is used to specify $\boldsymbol{H}(\boldsymbol{s})$ and $\kappa(\boldsymbol{s})$ in the non-stationary SPDE (28),

$$\boldsymbol{H}(\mathbf{s}) \equiv \exp(\theta_1)\boldsymbol{I}, \quad \kappa(\mathbf{s}) = \exp(\theta_2)\left[1 + \theta_3\psi(\mathbf{s})\right], \tag{31}$$

where $\psi(\mathbf{s})$ is obtained at every time step t by smoothing out the ensemble variance of $\boldsymbol{x}_t$ and normalizing the smoothed variance estimate so that $\max_{\mathbf{s}}|\psi(\mathbf{s})|$ is equal to 1. In practice this leads to basis functions with larger values near the shoreline, i.e. the intersection between the top layer and the sea surface, as this is where the largest variances in thickness are found. An alternative way to create such basis functions would be to compute the location of the shoreline explicitly, and calculate the distance from every grid cell to the closest point on the shoreline.

This is a parsimonious parametrization of the precision structure, with only 3 parameters. Tests with more complex basis functions either led to difficulties in the likelihood optimization, or yielded prediction results that were very similar to those obtained using the simpler parametrization. This is in line with the findings of Fuglstad et al. (2015b).

5.2. Filtering results

Fig. 9 shows estimates of the top surface $\mathbf{z}_{tt}$ at $t = 5$ and $t = 15$ for the three EnKF variants. While the estimated fields are rather similar between the three approaches, there are differences in standard deviation. Relative to the empirical approach, the parametric estimates have both a higher overall variance level, and a sharper transition between the high and low variability regions.

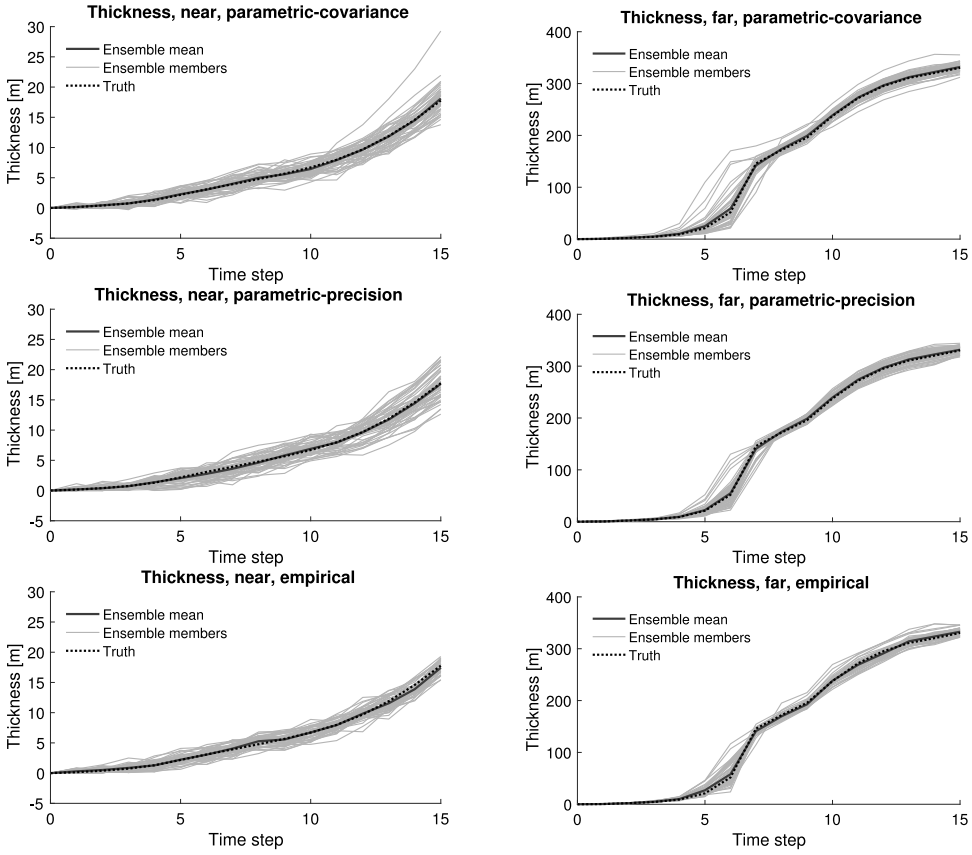

Fig. 10. Ensemble estimates of cumulative thickness x_t at time $t = 0, 1, \ldots, 15$ at two locations in block F3. Left: Thickness at location (15,5) which is close to Well 1. Right: Thickness at location (46,3) which in the southeast corner of the F3 block region, moderately far away from both wells.

Fig. 10 shows how the filtering ensemble evolves from $t = 1$ to $t = 15$ in each case. This display shows that the ensemble tracks the reference realization well in each case. The variability appears to be smaller at all times for the empirical approach.

Fig. 11 shows the location-wise rank of the reference thickness relative to EnKF ensemble thicknesses at time $t = 5$ and $t = 15$. The brighter a cell is, the larger the reference value is relative to the ensemble members at the location in question. Since the modeled field is spatially correlated, aggregating ranks over the domain may be misleading. Still, for a well-calibrated filter one expects a uniform distribution of ranks over the integers $\{1, 2, \ldots, B, B + 1\}$. In the present case, however, we find that all three filter variants are overdispersive, with almost no ranks below 15 or over 35, which are the limits of the color ranges in Fig. 11. The rank plots of the two parametric filters have similar patterns, with the highest ranks concentrated near the shoreline, in the region where the estimated standard deviation is largest. This indicates that both parametric EnKF variants overestimate the variance in the high-variability region. In the rank plots of the empirical filter we see a different pattern, featuring large, contiguous regions of over- and underestimation. Furthermore, the rank plot at $t = 15$ appears to retain some features of the earlier pattern. The rank patterns of the parametric filters at $t = 15$ do not show clear traces of earlier patterns.

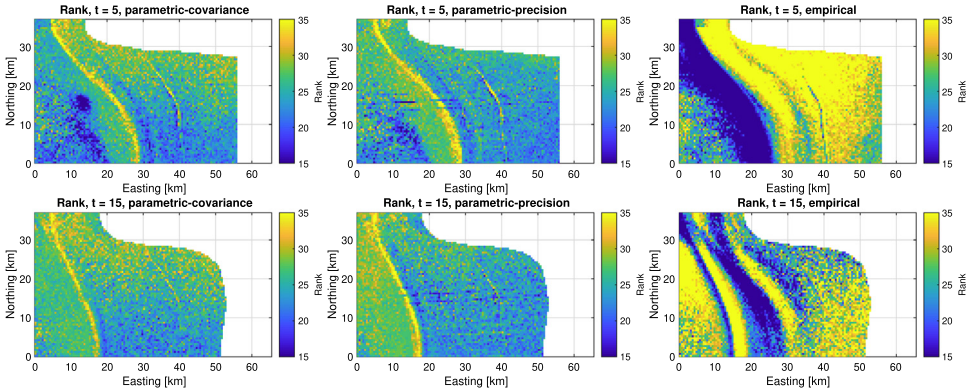

Fig. 11. Rank of reference realization total thickness relative to ensemble total thickness at $t = 5$ and $t = 15$ for parametric covariance, parametric precision and standard empirical EnKF variants. The strip of missing values along the northern and eastern boundaries is due to all ensemble members being equal on this region.

6. Closing remarks

In this paper we have suggested some approaches for integrating more spatial statistical modeling in ensemble-based filtering methods. Parametrizing the covariance used in the update step of the EnKF gave filter output with improved predictive performance. In particular, the underestimation of variance that is often seen in EnKF predictions was reduced.

The parametric models used include simple stationary covariance models, and semi-parametric models for the covariance or precision matrix structure. In judging the results, we noted that parsimonious covariance structures did surprisingly well.

Applying the EnKF with parametric covariance estimation to a data assimilation test case involving a geological process model (Section 5), we found little difference between the results of using an exponential covariance function specification and a GMRF precision matrix specification of the correlation structure of the target random field. However, comparing these parametric filters with a standard stochastic EnKF whose gain matrix is based on empirical sample covariance estimators, we found clearer differences. In our non-linear test case, going from empirical to parametric covariance estimation gave no obvious improvement in estimates of the overall level of variance, but did appear to produce a less systematic pattern of bias in the estimated random field.

While introducing parametric estimation into a larger workflow can improve the quality of estimates, we find that very simple parametrizations tend to be preferable to even slightly more complex ones, as the flexibility gained by adding an extra parameter seldom makes up for the added difficulty in estimation. Using non-stationary variance entries and single parameter correlation sometimes improved results, while having complex precision structures led to difficult likelihood surfaces, without always improving predictive power. This means that finding a useful parametric model to embed in the EnKF update can be relatively easy, as one can bet on simplicity by choosing an uncomplicated model. Even if increased flexibility gives a better description of the random field being modeled, it does not follow that the estimates obtained under the more flexible model will be better than ones obtained under the simpler model in terms of predictive ability.

None of the parametric approaches studied here included any kind of anisotropy. This extension would give a few extra model parameters to estimate, and could be interesting for some applications. We conducted maximum likelihood estimation separately at every time step. The procedure could be extended to have coupling of parameters at different time steps. One could also apply Bayesian hierarchical models in this context.

References

Asch, M., Bocquet, M., Nodet, M., 2016. Data Assimilation: Methods, Algorithms, and Applications. SIAM.

Asfaw, Z.G., Omre, H., 2016. Localized/shrinkage kriging predictors. Math. Geosci. 48 (5), 595–618.

Beaumont, M.A., 2010. Approximate bayesian computation in evolution and ecology. Annu. Rev. Ecol. Evol. Systematics 41, 379–406.

Charvin, K., Gallagher, K., Hampson, G.L., Labourdette, R., 2009. A bayesian approach to inverse modelling of stratigraphy, part 1: method. Basin Res. 21 (1), 5–25.

Cressie, N., Wikle, C.K., 2011. Statistics for Spatio-Temporal Data. John Wiley & Sons.

Doucet, A., Godsill, S., Andrieu, C., 2000. On sequential monte carlo sampling methods for bayesian filtering. Statist. Comput. 10 (3), 197–208.

Evensen, G., 2009a. Data Assimilation: the Ensemble Kalman Filter. Springer Science & Business Media.

Evensen, G., 2009b. The ensemble Kalman filter for combined state and parameter estimation. IEEE Control Syst. 29 (3), 83–104.

Fuglstad, G.A., Lindgren, F., Simpson, D., Rue, H., 2015a. Exploring a new class of non-stationary spatial Gaussian random fields with varying local anisotropy. Statist. Sinica 115–133.

Fuglstad, G.A., Simpson, D., Lindgren, F., Rue, H., 2015b. Does non-stationary spatial data always require non-stationary random fields?. Spat. Statist. 14, 505–531.

Furrer, R., Bengtsson, T., 2007. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. J. Multivariate Anal. 98 (2), 227–255.

Gneiting, T., Balabdaoui, F., Raftery, A.E., 2007. Probabilistic forecasts, calibration and sharpness. J. R. Stat. Soc. Ser. B Stat. Methodol. 69 (2), 243–268.

Hou, E., Lawrence, E., Hero, A.O., 2016. Penalized ensemble kalman filters for high dimensional non-linear systems, arXiv preprint arXiv:1610.00195.

Jun, M., Stein, M.L., 2008. Nonstationary covariance models for global data. Ann. Appl. Stat. 1271–1289.

Katzfuss, M., Stroud, J.R., Wikle, C.K., 2016. Understanding the ensemble Kalman filter. Amer. Statist. 70 (4), 350–357.

Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E., 1998. Convergence properties of the Nelder–Mead simplex method in low dimensions. SIAM J. Optim. 9 (1), 112–147.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B Stat. Methodol. 73 (4), 423–498.

Myrseth, I., Omre, H., 2010. Hierarchical ensemble Kalman filter. SPE J. 15 (02), 569–580.

Neto, J.H.V., Schmidt, A.M., Guttorp, P., 2014. Accounting for spatially varying directional effects in spatial covariance structures. J. R. Statist. Soc. Ser. C (Appl. Statist.) 63 (1), 103–122.

Ogg, J.G., Ogg, G.M., Gradstein, F.M., 2016. The Concise Geologic Time Scale. Elsevier.

Paciorek, C.J., Schervish, M.J., 2006. Spatial modelling using a new class of nonstationary covariance functions. Environmetrics 17 (5), 483–506.

Parker, R.J., Reich, B.J., Eidsvik, J., 2016. A fused lasso approach to nonstationary spatial covariance estimation. J. Agric. Biol. Environ. Statist. 21 (3), 569–587.

Parquer, M.N., Collon, P., Caumon, G., 2017. Reconstruction of channelized systems through a conditioned reverse migration method. Math. Geosci. 49 (8), 965–994.

Petersen, K.B., Pedersen, M.S., et al., 2008. The Matrix Cookbook, Vol. 7. Technical University of Denmark, p. 15.

Sætrom, J., Omre, H., 2011. Ensemble Kalman filtering with shrinkage regression techniques. Comput. Geosci. 15 (2), 271–292.

Sætrom, J., Omre, H., 2013. Uncertainty quantification in the ensemble Kalman filter. Scand. J. Stat. 40 (4), 868–885.

Skauvold, J., Eidsvik, J., 2018. Data assimilation for a geological process model using the ensemble Kalman filter. Basin Res. 30 (4), 730–745.

Snyder, C., Bengtsson, T., Bickel, P., Anderson, J., 2008. Obstacles to high-dimensional particle filtering. Mon. Weather Rev. 136 (12), 4629–4640.

Stroud, J.R., Katzfuss, M., Wikle, C.K., 2018. A bayesian adaptive ensemble kalman filter for sequential state and parameter estimation. Mon. Weather Rev. 146 (1), 373–386. http://dx.doi.org/10.1175/MWR-D-16-0427.1.

Tetzlaff, D.M., 2005. Modelling coastal sedimentation through geologic time. J. Coast. Res. 21 (3), 610–617.

Tsyrulnikov, M., Rakitko, A., 2017. A hierarchical Bayes ensemble Kalman filter. Physica D 338, 1–16.

Ueno, G., Higuchi, T., Kagimoto, T., Hirose, N., 2010. Maximum likelihood estimation of error covariances in ensemble-based filters and its application to a coupled atmosphere–ocean model. Q. J. R. Meteorol. Soc. 136 (650), 1316–1343.

Ueno, G., Nakamura, N., 2016. Bayesian estimation of the observation-error covariance matrix in ensemble-based filters. Q. J. R. Meteorol. Soc. 142 (698), 2055–2080.

Ueno, G., Tsuchiya, T., 2009. Covariance regularization in inverse space. Q. J. R. Meteorol. Soc. 135 (642), 1133–1156.

Wingate, D., Kane, J., Wolinsky, M., Sylvester, Z., 2016. A new approach for conditioning process-based geologic models to well data. Math. Geosci. 48 (4), 371–397.